

An Effective Pipeline for Pedestrian Detection in Mid-High Density Crowd

Yu Hao, Duke Gledhill, Ying Liu, Jiulun Fan, Zhijie Xu

School of Computer Science / School of Computing and Engineering
Xi'an University of Posts and Telecommunications / University of Huddersfield
Xi'an, China / Huddersfield, UK

Abstract—A novel framework is introduced to handle the pedestrian detection in mid-high crowd density. This framework exploits the samples obtained with the unsupervised approach, extracts the combined pattern of Histograms of Oriented Gradients (HOG) and Local Binary Pattern (LBP) for the training and detection. The motion information is utilized to adjust the confidence score of detect annotation, and the soft-NMS is used to handle the balance between the removing of redundant annotation and occlusion. The experiments indicate the proposed approach achieved a promising result compared to state-of-the-art trained with the benchmarking dataset.

Keywords—Pedestrian Detection; HOG; LBP; Soft-NMS; Mid-high Crowd Density

I. INTRODUCTION

In order to ensure public safety, CCTV cameras are widely installed in public areas such as plazas and train stations. The video stream is utilized for predicting or analyzing anomalies among the pedestrians. For most approaches, collecting the spatial position and count of objects/pedestrians in the scene is often the initial process. Non-personal-based and personal-based modeling approaches are exploited to achieve the purpose of pedestrian counting and detection. The non-personal-based modeling approaches extract global patterns and estimate the number of objects/pedestrians in the scene without modeling the individuals. For example, a typical approach is proposed in the research of Chan [1]. Pedestrians in the scene are firstly segmented and extracted from the background. Based on the assumption that the long-term relation of segmented pixels and pedestrian count matches the linear distribution, 3 different types of features are used for the training of a Gaussian Process (GP) model. In this GP model, the linear kernel is used to capture the long-term linear trend and RBF kernel is used to capture the local trend. Once the GP regression is completed, the pedestrian count with any size of segmented pixels could be estimated.

Approaches of personal-based modeling detection are more widely explored. For example, a typical approach is to detect objects using HOG features and machine learning tools such as SVM. The HOG patterns of pedestrians are extracted and fed to SVM as positive training samples along with negative ones. In the detection phase, image patches selected by a sliding window are sent to the trained SVM for classification. In the research

of Desai [2], objects are firstly detected using the HOG and SVM approach. The co-occurrence between detected objects from different classes is then modelled to optimize the performance with non-maximum suppression. Instead of modeling the entire object with HOG feature, the research of Pedro [3] proposed a method of training the latent SVM with the HOG feature extracted from the different components of the object for the detection. This research [3] also devised a margin-sensitive approach for the mining of Hard Negative samples of the training data. The research of Mikel [4] exploited global crowd density information to improve the detection performance of pedestrian detection. In its model, the confidence score of the detected pedestrian is compared with the crowd density at its location. When the difference is high, the detected pedestrian would suffer a significant confidence reduction. As a consequence, detections with top confidence scores exhibit higher accuracy than the conventional detection approach.

Different visual patterns are widely exploited in the state-of-the-art research of pedestrian detection. In the conventional approaches, patterns extracted from a sliding window are used to train the machine learning models for classification. The sliding window object detection approach was originally proposed in the research of Papageorgiou and Poggio [5]. In the approach, a pattern of Haar wavelets is fed to the support vector machine to achieve the detection. Various following works [6][7][8] have been conducted based on this concept. Three main types of feature are most frequently utilized, which include gradient-based features, shape-based features and motion information. For the gradient-based features, the most efficient pattern for object detection is the Histogram of Oriented Gradient (HOG) as claimed in [9]. The HOG feature is firstly introduced in the work of Dalal [10] which is inspired by Scale-Invariant Feature Transform (SIFT) [11]. Additional efforts [12] have also been made to improve the extraction efficiency of HOG. As for the shape-based features, patterns such as contour or boundary information are exploited. In these approaches, the shape-based patterns are extracted from the sliding window and matched with the pre-defined shape templates to achieve object detection. In the research of Tomoki [13], instead of using HOG, the co-occurrence matrix of HOG (CoHOG) is modelled as the descriptor for classification. The experimental result shows the miss-rate is 50% lower using the CoHOG than conventional HOG. Another approach is to detect the components using the local shape

descriptors, then achieve the detection by analyzing the spatial relationships between these components. For example, the edgelet feature in the research of Nevatia [14] and the shapelet feature in the research of Sabzmejdani [15]. For the motion information, patterns such as optical flow are used to amend the accuracy of the detection result. In the research of Dalal [16], the internal difference of optical flow is used to compensate the local motion of a detected pedestrian. And density field is utilized to fix the confidence score of the detected pedestrian using the HOG and SVM approach in the research of Mikel [4]. However, the performance of using motion information degrades drastically when the camera is not static. Therefore, implementation of motion information under the circumstance of a moving camera is still a challenge to be addressed.

In this research, we introduced an improved pedestrian detection approach based on HOG and LBP patterns. The rescoring process using motion information is exploited to optimize the confidence score of the SVM classification. The soft-NMS procedure is implemented to handle the widely existing occlusion problem in mid-high crowd density. The contents of this research distribute as follows; Section 2 introduces the general structure of the proposed pedestrian detection approach. Section 3 introduces the visual pattern exploited in the proposed approach. Section 4 introduces the devised confidence score adjust mechanism based on the motion magnitude and introduces the adapted enhanced NMS approach to handle the bounding box annotation of occluded pedestrians. Section 5 shows the experimental results with the proposed pedestrian annotation approach on PETS09 dataset. Section 6 gives the conclusion.

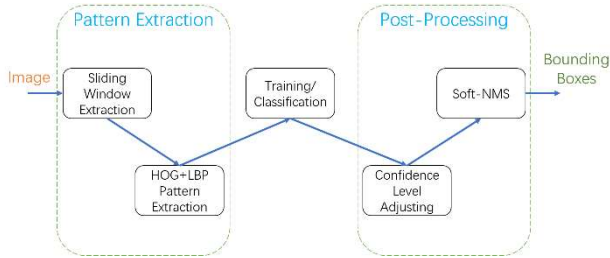


Figure 1. The general structure of proposed approach

II. STRUCTURE OF PROPOSED DETECTION APPROACH

As previously introduced, the procedure of personal-based modeling detection consists of three fundamental phases, which are pattern extraction, training/classification and post-processing. The structure of the proposed approach could be illustrated as Fig. 1. In the pattern extraction phase, the patches are extracted from the image with a sliding window. Before extracting the visual pattern from the image patch, a pre-processing approach is implemented to improve the time efficiency. In the pre-processing approach, a motion magnitude check is applied to the patch. If the check is failed, the corresponding patch will be discarded without pattern extraction. After the patch passes the motion magnitude check, the visual pattern will be modelled from these patches. In this approach, the HOG and LBP are adapted. The modelled pattern will be either sent to train the

classifier, or determined as a positive/negative sample assigned with a confidence score. In the post-processing phase, the obtained confidence score will be further adjusted based on the instant motion density to strengthen the accuracy. In order to handle the occlusion issue in the high-density situation, instead of using the conventional NMS process, an enhanced soft-NMS is adapted to remove the redundant bounding boxes.

III. THE HOG AND LBP BASED DESCRIPTOR FOR CLASSIFICATION

As claimed in the research of Piotr [9], by using single pattern only, the HOG achieved the best performance on object/pedestrian detection. In order to further improve the detection performance, the mixture of patterns is exploited. In the research of Wojek [17], patterns such as shapelets, shape context and HOG are used for detection and edgelet, HOG and covariance are exploited in the work of Wu [18]. In this research, the combined HOG and Linear Binary Pattern (LBP) are used as the primary patterns in the training and classification process which originated in the research of Wang [19]. Using conventional HOG pattern for object detection doesn't perform well when the static background contains intensive edges. On the opposite, the edge information such as LBP is capable of filtering these noise edges. As proved in [19], the combined pattern of HOG and LBP achieved better performance than using HOG only on the INRIA dataset [20]. Therefore, the HOG and LBP are adapted as the pattern for classification. The pattern modelling process is illustrated in Fig. 2.

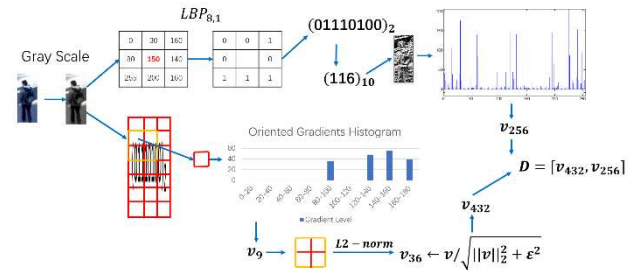


Figure 2. Modelling Procedure of the HOG and LBP pattern.

A. Descriptor Modelling

In order to model this pattern, the image patch from the training samples or sliding window is transformed into the gray scale format. Next, patterns of LBP and HOG are extracted from the patch separately. For the LBP pattern, $LBP_{n,r}$ indicates the pattern is extracted with n pixels around the target pixel with radius r . The relationship matrix is then converted into a binary pattern value then transformed into decimal format. This decimal value is the LBP of this pixel. As illustrated in Fig. 2, $LBP_{8,1}$ will derive a binary value with 8 digits, which is a decimal value with the range from 0-255 correspondingly. Therefore, a histogram is modelled on LBP of all pixels in the image patch. And a vector pattern v_{256} with 256-dimensions could be obtained from this histogram.

For the HOG pattern, the image is divided into multiple cells, as shown in Fig. 2. For each cell, a histogram of accumulative gradient magnitude in 9 orientation bins is modelled, and a 9-dimensional vector is

derived from the histogram. A block could be composed with n cells, 4 cells in the figure example. Then vectors of these n cells will form a combined vector which is unified using the L2-norm. By sliding blocks on the patch, a combined vector v_{432} with high dimension will be generated.

Once the LBP and HOG patterns are obtained, a 688-dimensional descriptor D is modelled for the training and classification process as (1).

$$D = [v_{432}, v_{256}] \quad (1)$$

B. Patch filtering preprocessing

The extraction of HOG and LBP patterns from an image patch is usually time-consuming. The image patch is obtained using the sliding windows. Thus, hundreds of patches could potentially be generated. If every patch needs to be analyzed, the overall processing time could be significantly long, which may greatly impact the efficiency of the detection approach.

For pedestrian detection based on video data, the instant motion within the footage could be obtained from two consecutive frames. The motion information could be exploited to estimate if the patch deserves to be further analyzed. In the proposed approach, instead of extracting patterns from every obtained patch, a preprocessing mechanism is implemented based on the magnitude of the calculated optical flow. If the patch failed the motion magnitude check, it will be discarded without the pattern extraction process. The motion magnitude check of patch P could be expressed as (2)

$$\frac{\sum_{i=1, j=1}^{M, N} ((M(P_{i,j}) > T) ? 1 : 0)}{M \times N} > 0.5 \quad (2)$$

Where M, N are the width and height of the image patch. M is the magnitude of optical flow at current location. In this research, the HS optical flow is adapted for the extraction. T is a threshold to measure the flow magnitude. If the flow on current pixel $M(P_{i,j})$ is larger than T , this pixel will be considered active. Overall, if the active pixels occupy over half of the entire patch, the motion magnitude check is successful. Otherwise, the check is failed and the patch will not be used for pattern extraction.

The implementation of this mechanism can greatly reduce the processing time of the pattern extraction phase. In the example of Fig. 4, over 70% patches are filtered. Since the pattern extraction is often the most time consumptive phase among the detection procedure, the overall processing time could also be significantly shortened.

IV. POST-PROCESSING

The modelled descriptor is fed to the trained classifier to determine whether the corresponding image patch contains a pedestrian. If the classification result is positive, the post-processing will be implemented to increase the

detection accuracy and handle the occlusion problem according to the introduced framework. In this section, the confidence score adjusting mechanism and adaption of soft-NMS will be explained.

C. Confidence score adjusting

When complete, the classifier will generate a set of bounding boxes to annotate the possible detections. Each bounding box is assigned as confidence score S , which represents the probability of the optimal detection. However, these scores don't always match the ground truth. For example, in the high crowd density, the frequent occlusion will severely hamper the accuracy of the score estimation. In order to further optimize the detection result, the information of local motion density is exploited to adjust the confidence score. Also, since the overlapped bounding boxes are generated from the previous process, the NMS process will be applied to reduce the redundancy of bounding boxes.

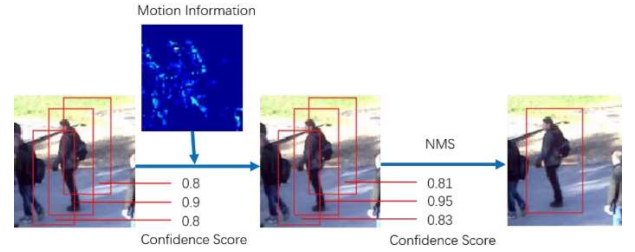


Figure 3. Confidence score adjusting

Inspired by the research of Mikel [4], the local crowd density is exploited to optimize the confidence scores. For each detection x , the density is estimated with a gaussian kernel around its neighbours, note as Ax . Next, the difference between local density D and Ax is used to adjust the confidence score. Inspired by it, the instant motion information is utilized for the optimization of the confidence score instead of using crowd density. The updating process of confidence score S'_x of any detection x could be expressed as (3).

$$S'_x = S_x + \alpha \text{AVG} \left(\sum_{j \in A} M(j) \frac{1}{\sqrt{2\pi}\sigma} \exp \left(-\frac{\|i-j\|^2}{2\sigma^2} \right) \right) \quad (3)$$

Which 2-dimensional array A contains all indices of an instant motion magnitude array M , and its value range is unified between 0 to 1. The size of M is determined by the size of bounding box. The value of i indicates the centered index of M . Therefore, every motion magnitude in $M(j)$ is reevaluated according to the gaussian function which measure the distance between i and j . When j is far from the center i , the corresponding magnitude will be greatly reduced and vice-versa. σ is set to the width of M . The α is the impact factor which decides how greatly the confidence score will be influenced.

The procedure of this section is illustrated as Fig. 3. Firstly, three patches are classified as pedestrian with different confidence score. Note the bounding box in the middle has the highest S 0.9, which indicates this bounding box has higher probability to be the optimal detection than the two adjacent boxes. Next, the instant

motion information is extracted from the two consecutive frames with HS or LK optical flow algorithm. By analyzing the motion density in the corresponding bounding box, its S will be adjusted. In this case, the S of the middle box has the largest increment 0.5, because the motion density in this box is highest among these three boxes.

D. Bounding Box redundancy removal

For the sliding window detection approach, the detection process will generate multiple overlapped bounding boxes with different confidence scores for the same object. When the step length of the sliding window is small, the number of redundant bounding boxes will be significantly high, as illustrated in Fig. 4(a). In order to address the redundancy of bounding boxes, the Non-Maximum Suppression (NMS) approach [2] should be applied. After the process of NMS, the overlapped bounding boxes are removed according to their confidence score, as illustrated in Fig. 4(b).



Figure 4. (a) Redundant bounding boxes. (b) NMS processed result

The NMS is widely utilized in the research of object detection. The fundamental concept of NMS is to suppress the elements with non-maximum value, which could be categorized as a local maximum search. After the patterns of the sliding windows are recognized with the classifiers, each bounding box will get a probability score. Then the NMS is applied to obtain the local maximum probability score of overlapped bounding boxes. The detailed procedure of conventional NMS is as follows. For the list of bounding boxes B and the confidence score S , move the bounding box M with highest S from B into the result set D , and remove any other bounding boxes with the Intersection-Over-Union (IOU) to M larger than a threshold N_t from B . Repeat this procedure until B is empty. For the multi-class detection, the NMS procedure is applied for every class. However, when a target is detected as two or more different classes and the shared IOU is large, the detection result will be unacceptable, i.e. one object is categorized as multiple objects with different classes. In order to address this issue in multi-class NMS, the so-called multi-task loss for NMS score is devised to suppress the probability between two different classes in the research of Liu [22]. With the proposed multi-task loss, S of an object with high IOU between multiple classes will be significantly reduced, and the target will be successfully detected as a single object.

Another critical issue of the above-mentioned approaches is when two objects of same class are heavily occluded, the NMS tends to remove the bounding box with smaller S which makes the occluded object fail to be detected. In the research of Bodla [23], the Soft-NMS is introduced to address this problem. Instead of removing the bounding box with smaller S and $IOU > N_t$, its S is

assigned with an updated value calculated either from a linear or gaussian function of current S . It claims a 1% improvement when applying Soft-NMS to the testing phase of R-FCN and Faster-RCNN models on MS-COCO dataset. The main aim of this research is to detect the pedestrians in the crowd state. Therefore, occlusion between pedestrians will frequently occur. In this research, the Soft-NMS is applied to address the occlusion issue. There are two penalty function to adjust the overlapped bounding box's S , which are linear and gaussian functions illustrated in (4) and (5).

$$S = \begin{cases} S, & IOU < N_t \\ S(1 - IOU), & IOU \geq N_t \end{cases} \quad (4)$$

Similar to the conventional NMS, for each bounding box M whose IOU is smaller than the N_t , its confidence score S remains the same. Otherwise, its S should be suppressed according to the IOU ranged from 0 to 1. However, the linear penalty function isn't continuous. When the value of IOU reaches N_t , the value of S would experience a sudden drop. This may cause a drastic changing of the confidence list. With the gaussian penalty function, when the intersection area is small, the impact of penalty should be small and gradual. On the contrary, if the intersection is large, the value of S should be greatly reduced. In order to achieve a continuous trend of S , the gaussian function is exploited as below.

$$S = S \exp\left(-\frac{(IOU)^2}{\sigma}\right) \quad (5)$$

Once all bounding boxes' confidence scores are rescored with the penalty function, a threshold N_s is used to prune boxes with smaller S .

V. EXPERIMENT RESULTS

In this section, performance of the proposed approach on pedestrian detection in high density will be investigated through three different experiment installations. In the first experiment, the performance between typical HOG and HOG+LBP on PETS09 dataset is compared. In the second experiment, the performance of the detection with Confidence Score Adjusting and soft-NMS is compared to the detection without the adapted mechanism. The result indicates the proposed detection approach outperformed the conventional one on the crowd with mid-high density.

A. Dataset for experiment

Two video datasets will be exploited to inspect the performance of proposed detection approach. The first one is the PETS2009 dataset [21]. This set contains various videos simultaneously recorded with multiple static cameras installed at different locations. In each video, a group of pedestrians with high density walks as a crowd. Thus, this dataset satisfies the crowd density requirement of the proposed detection approach. The PETS2009 contains three sub-sets for different purposes including pedestrian counting, tracking and event recognition. In the experiment, the pedestrian counting subset is primarily

utilized. In this subset, pedestrians usually appear within a single group in mid-to-high density. The crowd moves from one side of the footage to the other with four view-perspectives. These patterns make the PETS2009 to be the most proper dataset for the proposed pedestrian detection approach in high crowd density.

The INRIA dataset will also be exploited in the experiment [10]. This set is used as a benchmark for the individual detection and widely adopted in research. It contains thousands of image data which label the information for training and testing such as positive/negative, co-ordinates and ground truth. However, most footage in the INRIA contains sparse individuals, which isn't suitable for testing the detection in high density. Also, the proposed approach is devised to handle the detection in the video sequence from a single camera. However, the INRIA set contains inconsecutive images from different sources. Therefore, only a portion of INRIA set will be adopted for the training purpose in the experiment.

B. Comparison between typical HOG and HOG+LBP

In the first experiment, the detection results of typical HOG and HOG+LBP are compared to verify the later one has better performance. Fig. 5 illustrates the pedestrian detection results using the unsupervised training approach in previously proposed research [13] on HOG pattern only and the devised descriptor with HOG and LBP. Comparing with the Fig. 5(b), it is obvious that more false positive detections occurred in Fig. 5(a), and result in Fig. 5(b) achieved higher recall rate. In order to further prove the performance of using the combined pattern of LBP and HOG, the False Positive Per Image – Miss Rate (FPPI-MR) curve is modelled by applying the proposed approach with these two patterns on the PET 2009. The supervised training uses subset S1/L2/Time_14_31, and the video data for detection uses subset S1/L2/Time_14-06. The overlap threshold of the FPPI and MR checking is set to 50%. These two subsets contain videos from the same static camera, but with different lighting conditions and crowd behaviour. The result implies that the HOG+LBP pattern outperformed HOG only with 10% less miss-rate on the same FPPI. Compared to the implementation on NUS dataset in [13], the HOG+LBP pattern is proved to have a better performance on occluded crowd as well.

C. Comparison between proposed and conventional on PETS2009

In this experiment, the performances are evaluated on different combinations of patterns using the same unsupervised sample extraction approach. In the training phase, data from subset S1-L2-Time_14-06 of each camera is used for positive and negative sampling. This subset contains 200 consecutive frames, all of which will be feed to the unsupervised sampling model. The size of image patch is set to 18 by 18 pixels. The calculated average cluster size is 10 for a total 200 frames. In the detection phase, subset S1-L2-Time_14-31 is used for the detection. This subset contains 130 frames which have the identical view perspective but different lightness to the training data. According to the estimated pedestrian size, the bounding box is empirically set as 100 by 40 pixels.

The linear SVM is trained and applied for the detection. The step of sliding window is set to 5 pixels. In the rescoring phase, impact factor α is set to 0.2. In the soft-NMS phase, the threshold is set to 0.1.

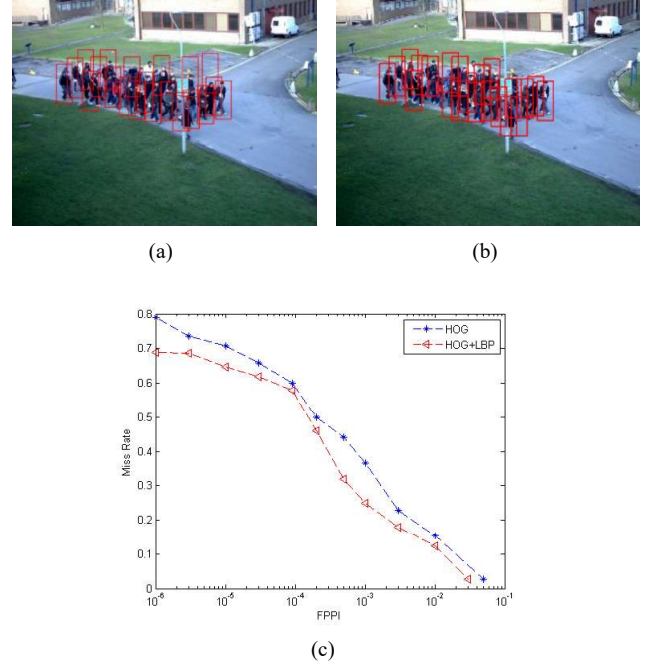


Figure 5. (a) Detection result with HOG only. (b) Detection result with HOG and LBP. (c) The MR-FPPI curve using HOG+LBP and HOG

The detection result is illustrated in Fig. 6, as long as the results of applying the combination of standard patterns. The HOG+LBP+MR+SNMS line indicates the detection result using HOG and LBP patterns along with Motion Rescoring (MR) and Soft-NMS (SNMS). The HOG+LBP line indicates the result with ordinary NMS. The HOG line indicates the result with HOG and ordinary NMS. A linear SVM is applied on all approaches. The result indicates the proposed approach achieved better performance than other approaches on this subset of PETS09 (Lower is better).

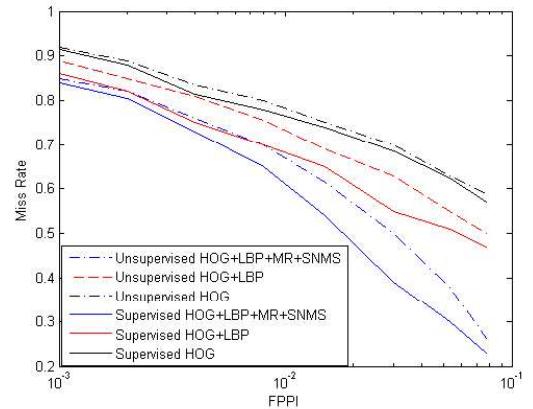


Figure 6. Comparison between the detection results with and without adapting the adapted post-processing approach

VI. CONCLUSION AND FURTHER DISCUSSION

In this research, a novel pedestrian detection framework is devised to handle the detection in the crowd with mid-high density. For the pattern extraction process, this framework adapts the HOG and LBP patterns for the training and testing of the classifier. After the bounding boxes are detected, a confidence score adjusting mechanism is devised based on the instant local motion to improve the accuracy of NMS process. In order to handle the heavy occlusion in high crowd density, the soft-NMS with the gaussian penalty function is adapted instead of the conventional NMS.

In the experiment, PETS2009 dataset is exploited to inspect the performance of the proposed detection approach. The FPPI-MR curve indicates the enhanced approach of using HOG+LBP pattern improves the performance of the conventional HOG pattern. The experiment also indicates the confidence rescore adjusting mechanism and soft-NMS improves the detection accuracy on PETS2009 dataset. However, the adapting of motion information in pre-processing and confidence score adjusting makes the approach achieve better performance on video data than static images. If the data source is a static image, the efficiency in pre-processing would be hampered. The confidence score adjusting mechanism doesn't apply on the static image as well. Future enhancement of the proposed approach would concentrate on the adaptiveness of the static images without losing the efficiency.

ACKNOWLEDGEMENT

This research is sponsored by Chinese National Natural Science Foundation (No. 61671377), and Key Research and Development Program of Shaanxi (2019GY-054).

REFERENCES

- [1]. B. Antoni, C. Zhang-Sheng, L. John, and N. Vasconcelos. "Privacy preserving crowd monitoring: counting people without people models or tracking". 2008 IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, USA. June 2008.
- [2]. C. Desai, D. Ramanan, and Fowlkes. "Discriminative models for multi-class object layout". International Journal of Computer Vision. Vol. 95, Issue 1, pp. 1-12, October 2011.
- [3]. P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. "Object detection with discriminatively trained part-based models". 2010 IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. 32, Issue 9, pp. 1627-1645, September 2010.
- [4]. M. Rodriguez, I. Laptev, J. Sivic, and J. Y. Audibert. "Density-aware person detection and tracking in crowds". 2011 IEEE International Conference on Computer Vision (ICCV 2011). Barcelona, Spain. November 2011.
- [5]. C. Papageorgiou, and T. Poggio. "A trainable system for object detection". International Journal of Computer Vision. Vol. 38, Issue 1, pp. 15-33, June 2000.
- [6]. P. A. Viola, and M. J. Jones. "Robust real-time face detection". International Journal of Computer Vision. Vol. 57, Issue 2, pp. 137-154, May 2004.
- [7]. A. Shashua, Y. Gdalyahu, and G. Hayun. "Pedestrian detection for driving assistance systems: single-frame classification and system level performance". 2004 IEEE Intelligent Vehicles Symposium. Parma, Italy. June 2004.
- [8]. S. Maji, A. Berg, and J. Malik. "Classification using intersection kernel SVMs is efficient". 2008 IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, USA. June 2008.
- [9]. P. Dollár, C. Wojek, B. Schiele, and P. Perona. "Pedestrian detection: an evaluation of the state of the art". IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. 34, Issue 4, pp. 743-761, April 2012.
- [10]. N. Dalal and B. Triggs. "Histograms of oriented gradients for human detection". 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). San Diego, USA. June 2005.
- [11]. D. G. Lowe. "Distinctive image features from scale-invariant key points". International Journal of Computer Vision. Vol. 60, Issue 2, pp. 91-110, November 2004.
- [12]. F. M. Porikli. "Integral histogram: a fast way to extract histograms in cartesian spaces". 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). San Diego, USA. June 2005.
- [13]. H. Yu, D. Maples, Y. Liu, and Z. Xu. "Unsupervised pedestrian sample extraction for model training". 12th IADIS International Conference on Information Systems 2019. Utrecht, Netherlands. April 2019. To be published.
- [14]. B. Wu and R. Nevatia. "Detection of multiple, partially occluded humans in a single image by bayesian combination of edgelet part detectors". Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1. Beijing, China. October 2005.
- [15]. P. Sabzmeydani and G. Mori. "Detecting pedestrians by learning shapelet features". 2007 IEEE Conference on Computer Vision and Pattern Recognition. Minneapolis, USA. June 2007.
- [16]. N. Dalal, B. Triggs, and C. Schmid. "Human detection using oriented histograms of flow and appearance". 2006 European Conference on Computer Vision (ECCV 2006). Graz, Austria. May 2006.
- [17]. C. Wojek and B. Schiele. "A performance evaluation of single and multi-feature people detection". 30th annual pattern recognition symposium. Munich, Germany. June 2008.
- [18]. B. Wu and R. Nevatia. "Optimizing discrimination-efficiency tradeoff in integrating heterogeneous local features for object detection". 2008 IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, USA. June 2008.
- [19]. W. Xiaoyu, H. X. Tony, and Y. Shuicheng. "An HOG-LBP human detector with partial occlusion handling". 2009 IEEE 12th International Conference on Computer Vision (ICCV 2009). Kyoto, Japan. October 2009.
- [20]. N. Dalal and B. Triggs. "Histograms of oriented gradients for human detection". 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). San Diego, USA. June 2005.
- [21]. PETS 2009 Benchmark Data. Online. <http://www.cvg.reading.ac.uk/PETS2009/a.html>
- [22]. L. Zikun, H. Jingao, W. Lubin, and Y. Yiping. "Rotated region-based CNN for ship detection". 2017 IEEE International Conference on Image Processing (ICIP). Beijing, China, September 2017.
- [23]. B. Navaneeth B, S. Bharat, C. Rama, and D. S. Larry. "Improving object detection with one line of code". International Conference on Computer Vision 2017. Venice, Italy. October 2017.